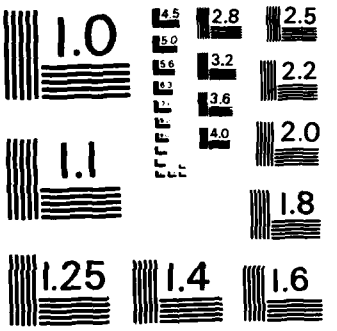
AD-A162 377 USE OF DATA ENVELOPMENT ANALYSIS IN COST ANALYSIS AND ESTIMATING(U) AIR FORCE INST OF TECH WRIGHT-PATTERSON AFB OH SCHOOL OF SYSTEMS AND LOGISTICS W F BOWLIN 1/1

UNCLASSIFIED SEP 85 F/G 14/1 NL



MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS - 1963 - A

(1)

Use Of Data Envelopment Analysis In Cost Analysis And Estimating

by

Major William F. Bowlin
Department of Systems Acquisition Management
School of Systems and Logistics
Air Force Institute of Technology

Presented at the
19th Annual Department of Defense
Cost Analysis Symposium
Xerox Training Center
Leesburg, Virginia
September 17 - 20 1985

AD-A162 377

DTIC FILE COPY

DTIC
ELECTE
NOV 26 1985
S D
B

DISTRIBUTION STATEMENT A
Approved for public release;
Distribution Unlimited

USE OF DATA ENVELOPMENT ANALYSIS IN COST ANALYSIS AND ESTIMATING

by

Major William F. Bowlin
Department of Systems Acquisition Management
School of Systems and Logistics
Air Force Institute of Technology

Introduction

Recently a new analytical technique called Data Envelopment Analysis (DEA) has appeared in the literature. The formulators of this methodology, A. Charnes, W.W. Cooper, and E. Rhodes [7], developed it with the idea of evaluating the performance (measuring productivity or efficiency) of not-for-profit organizations. However, it appears that DEA also has potential as a tool for use in "traditional" cost estimating/analysis roles.

The purpose of this paper is to briefly introduce the DEA methodology to the cost analysis community. We proceed in this paper by first presenting the DEA model formulation. This is followed by a description of the characteristics and conventions of the DEA model. The next section provides an example of the formulation of the DEA model for a specific analysis. We conclude the paper with a discussion on possible avenues of DEA use in cost estimating/analysis.

Data Envelopment Analysis Model

Fractional Model

DEA computes a relative measure of efficiency represented by h_0^* which is a scalar value. This measure is calculated as a ratio of weighted outputs to weighted inputs where the weights are determined by the model to maximize the value of this ratio for the decision making unit (DMU)[1] being evaluated. This maximization is subject to the condition that the corresponding values of the ratios for every organization used as a basis of comparison (including the unit being evaluated) must be positive and less than or equal to one. Mathematically this takes the form of the following model.

$$\text{Maximize:} \quad h_0 = \frac{\sum_{r=1}^{s} u_r y_{r0}}{\sum_{i=1}^{m} v_i x_{i0}} \tag{1}$$

$$\text{Subject to :} \quad 1 \geq \frac{\sum_{r=1}^{s} u_r y_{rj}}{\sum_{i=1}^{m} v_i x_{ij}}$$

$$j = 1,\ldots,n$$

$$u_r,\ v_i \geq e > 0$$

where the terms represent:

h_0 = The measure of efficiency for DMU "0", the member of the set of j = 1,...,n DMUs that is to rated relative to the others. The ratio on which h_0 depends is represented in the functional for optimization as well as in the constraints. This DMU preserves its original subscript identification in the constraints but is distinguished by a "0" subscript in the functional.

u_r = The variable for each type of output "r", which will be optimally determined by the solution of the model and assigned as a weight to the observed output value, y_r.

v_i = The variable for each type of input "i", which will be optimally determined by the solution of the model and assigned as a weight to the observed input value, x_i.

y_{r0} = The known amount of output "r" produced by DMU "0" during the period of observation.

x_{i0} = The known amount of input "i" used by DMU "0" during the period of observation.

y_{rj} = The known amount of output "r" produced by DMU "j" during the period of observation.

x_{ij} = The known amount of input "i" used by DMU "j" during the period of observation.

$e > 0$ = A small "non-Archimedean" constant.

Execution of the model requires repeated computations which, in principle, must be done for each DMU in the universe of organizations under evaluation.

The DEA model computes a rating of one for the organization under evaluation (meaning the organization is efficient) if, when compared to the other organizations in the set of organizations (including itself) being evaluated, it is operating efficiently.

Reference to Figure 1 on page 4 will help to show what is involved in the execution of this model.

QUALITY INSPECTED 3

Accession For
PER FORM 50
Availability Codes
Dist Special
A-1

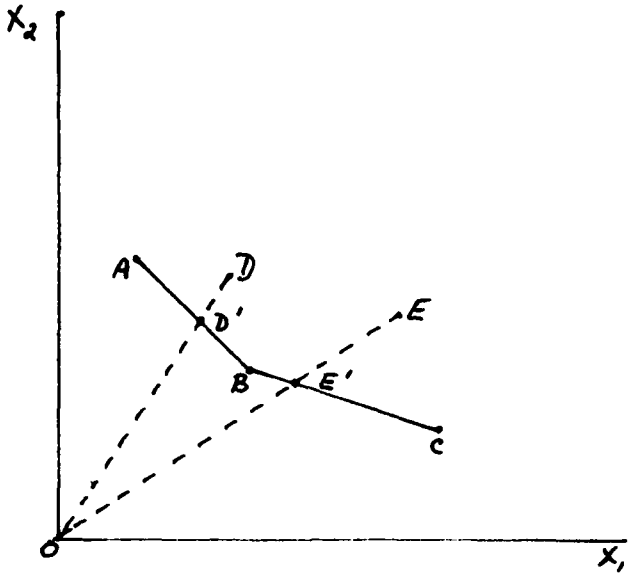


FIGURE 1 - Two Input/Single Output Case

The solid line connecting points A, B, and C represent a section of the unit isoquant, i.e., the level of the production surface, for one unit of a single output. For simplicity we restrict ourselves to the case of one output (produced at unit level) and two inputs, x_1 and x_2. The x_1 and x_2 coordinates of points A, B, C, D, and E represent observed inputs used to produce the one unit of output attained by each of the five DMUs associated with these points.

Both D and E are inefficient since they are dominated by D' and E' respectively and thus $h_0^{*4} < 1$ for both of these DMUs. The latter are not actually observed values but are obtained as convex combinations of A and B and B and C, respectively, which represent elements of the efficient frontier production possibility set. In fact, the values of h_0^* for points D and E correspond to the ratios of the ray segments d(O-D')/d(O-D) and d(O-E')/d(O-E) which are clearly less than unity.[5]

The points A, B, and C from which these convex combinations are obtained are all efficient and form an efficiency frontier. There is no point that can be generated from convex combinations of members of the production possibility set that will dominate them. Additionally, any movement along this frontier requires tradeoffs between x_1 and x_2 in order to stay on the frontier.

Linear Programming Model

The model previously presented is a non-linear programming problem. It is, in fact, a fractional programming problem with a linear fractional objective and linear fractional constraints. As such it is both nonlinear and nonconvex. However, Charnes, Cooper, and Rhodes [7] have shown that it can be transformed into an equivalent linear programming problem by means of the theory of linear fractional programming, developed by Charnes and Cooper [6]. The linear programming form offers many advantages such as dual variables and is the form commonly used for analysis. In order to simplify matters we are bypassing the development of the linear programming form and present it as follows:

Minimize: $h_0 = T - e\left(\sum_{r=1}^{s} s_r^+ + \sum_{i=1}^{m} s_i^-\right)$

(2)

Subject to:

$$\sum_{j=1}^{n} y_{rj} b_j - s_r^+ = y_{r0};\ r = 1,\ldots,s$$

$$-\sum_{j=1}^{n} x_{ij} b_j - s_i^- + T x_{i0} = 0;\ i = ,\ldots,m$$

$$b_j,\ s_r^+,\ s_i^- \geq 0$$

T unrestricted in sign

where:

T = An intensity value or multiplier of the observed input x_{i0}.

s_r^+ = Output slack for output "r".

s_i^- = Input slack for input "i".

e = A small positive valued non-Archimedean constant.

b_j = A variable whose value is determined in the solution of the DEA model.

DEA Extensions

Models (1) and (2) are basic formulations for data envelopment analysis and provide a measure of efficiency which combines technical and scale efficiencies. They also assume that all input and output variables are controllable. Since the initial development of DEA in 1978 there have been several extensions of the model which correct these deficiencies and other limitations and thereby make it more useful to managers.

Banker, Charnes, and Cooper [1] have modified the original model so that technical and scale inefficiencies can be separately identified. This allows management to review the technical efficiency of its operation at its current scale of operations.

Banker and Morey [2 and 3] have extended the model in two ways. First, they have modified the DEA model so that discretionary and nondiscretionary variables can be accommodated. This modified DEA model computes the efficiency of the organization while holding the noncontrollable variables constant. This allows management to take action on those variables that are controllable yet still consider essential noncontrollable variables in the model. Banker and Morey's second extension deals with handling categorical variables and allows this type of variable to be included in the efficiency analysis. In the original formulation of DEA these types of variables could not be included in the analysis.

The most recent extension is one by Charnes, Cooper, and Thrall [9]. They have shown how to accommodate decision making units that have some variables with zero values. However, this extension still needs to be operationalized.

Characteristics and Conventions of DEA Model

Characteristics

In this section we discuss some of the characteristics of the DEA model, and in doing so compare some of them to common regression approaches in order to clarify them.

First, DEA is an extremal or optimization methodology which optimizes on each observation (= DMU). By optimizing on each individual point, DEA establishes an efficiency frontier based on optimum performance. This contrasts with common regression

approaches which are "averaging" techniques and "average" over all observations - including efficient and inefficient organizations. Figure 2 illustrates this difference via a single output (Y), single input (X) case.

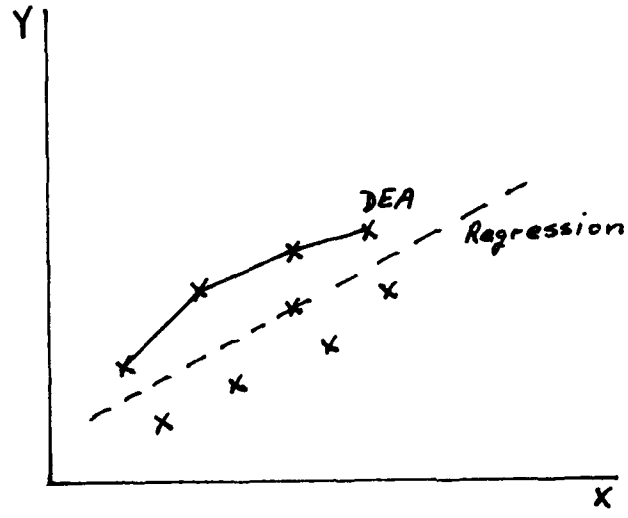


FIGURE 2 - DEA vs. Regression

This optimization principle of DEA allows for isolation on individual decision making units which is ideal for individual organizational control and permits the identification of sources and amounts of inefficiency or excess costs for each DMU. Knowing the sources and amounts of inefficiencies allows us to compute the level of input (costs) that should have been consumed (incurred) for a specific DMU via the following equations from Charnes, Cooper, and Rhodes [7]:

(3) $\hat{y}_r = y_r + s_r^{+*}$

(4) $\hat{x}_i = h_0^* x_i - s_r^{-*}$

where:

$\hat{y}_r$ = The efficient output level for output "r".

$\hat{x}_i$ = The efficient input level for input "i".

y_r, s_r^{+*}, h_0^*, x_i, s_i^{-*} are as previously described.

The DEA optimization principle also allows us to track the sources of comparison or reference set for the evaluated DMU via the b_j values in model (2). A positive b_j value indicates that DMU "j" is a member of the reference set for the DMU being evaluated.

Regression, on the other hand, does not offer such information for the manager to act upon, but simply indicates where an organization should be compared to all observations.

Second, the DEA model is a deterministic model that provides relative evaluations by creating an "efficient frontier" generated from actual observations. It is relative in the sense that the efficiency rating depends on the DMUs used. However, it does not depend on prior theoretical knowledge or explicit assumptions about the production process as in the model specifications used in statistical regression.

Third, DEA is able to handle multiple inputs and multiple outputs simultaneously and thereby recognizes the interaction or effect the multiple inputs and multiple outputs can have on each other. This contrasts with common regression models which can handle only a single dependent variable and multiple independent variables.

Fourth, the resulting efficiency value, h_0^*, does not depend on the units of measure in which the inputs and outputs are stated. That is, if any input or output is measured in different units then the value of h_0^* will not alter provided this same change is made in the units of measure for every other DMU in the comparison set.

Conventions

Recall that DEA is a measure of efficiency relative to the other organizations included in the evaluation. Thus all organizations in the reference set must be similar. That is, all organizations are assumed to have common inputs and outputs.

Second, all input and output values should have a positive value, and it is assumed an increase in input will result in an increase in output.

Third, a rule of thumb for maintaining an adequate number of degrees of freedom when using DEA is to obtain at least two DMUs for each input or output measure. Note, for instance, that an insufficient number of DMUs for the variables being used, would tend to produce a result in which all of the DMUs would be rated as 100% efficient simply because of an inadequate number of degrees of freedom.

Finally, there are two ways to accommodate the "e" in the DEA linear programming model. One way is to assign a very small value such as .000001 to it and then scale all variables to a value between 0 and 100. This scaling of variables is necessary so that the choice of T* by the model is not influenced by the slack values. An alternative approach is to solve the DEA model

in a two-step process. First, solve the model with the objective function being: minimize T. In the second step maximize the slack values in the objective function while constraining T to its value from step 1.

DEA Illustration

The following problem illustrates the formulation and implementation of the DEA model. The problem was taken from Bowlin [4] who analyzed the efficiency of real property maintenance activities at nine Air Force bases. The input and output measures used in this analysis are briefly described in Table 1 and the annual observed values of these variables are shown in Table 2. Executing model 2 for each of the nine bases provides an efficiency measure and variance analysis for each of these bases.

[Place Table 1 here.]

[Place Table 2 here.]

Model (5) below illustrates the formulation of a DEA problem by using the data from Table 2 for Base B in model (2). The problem for Base B would be set up as follows:

$$\text{Minimize: } h_0 = T - e(s^+_{CWO} + s^+_{CJO} + s^+_{CRWA} + s^+_{DJO} + s^-_{VEH} + s^-_{DOL} + s^-_{LABHR})$$

(5)

Subject to:

(5A) $197b_A + 135b_B + 162b_C + \ldots + 171b_I - s^+_{CWO} = 135$

(5B) $16,878b_A + 30,130b_B + \ldots + 10,773b_I - s^+_{CJO} = 30,130$

(5C) $12,860b_A + 3,492b_B + \ldots + 8,453b_I - s^+_{CRWA} = 3,492$

(5D) $.00042b_A + .0002b_B + \ldots + .00052b_I - s^+_{DJO} = .0002$

(5E) $40T - 44b_A - 42b_B - 105b_C - \ldots - 38b_I - s^-_{VEH} = 0$

(5F) $2,887.3T - 2,444.7b_A - 2,887.3b_B - \ldots - 2,406.0b_I - s^-_{DJO} = 0$

(5G) $338,611T - 265,866b_A - 338,611b_B - \ldots - 237,951b_I - s^-_{LABHR} = 0$

(5H) $b_A, \ldots, b_I, s^+_{CWO}, s^+_{CJO}, s^+_{CRWA}, s^+_{DJO}, s^-_{VEH}, s^-_{DOL}, s^-_{LABHR} \geq 0$

A few notes about the above model. First, each base is represented in each constraint and there is a constraint for each input and output variable. The input/output variables that the constraints apply to are indicated by the subscript of the slack variable, s. Constraints (5A) through (5D) apply to the output variables, (5E) through (5G) to the input variables, and (5H) reflects the nonnegativity requirements. For example, constraint (5A) is the constraint dealing with completed work orders (CWO) and each base is represented in the equation as indicated by the subscripted b variable where the subscript indicates the base.

Second, note that the values used for the delinquent job orders (DJO) output variable (equation 5D) is the reciprocal of the values shown in Table 2. This is because DJOs are not

desirable and therefore to meet the assumption of the DEA model that an increase in an input results in an increase in an output the reciprocal is used.

Third, we only had nine DMUs which violates the rule of thumb that there should be two DMUs for each input/output variable used in the model. For our model there should have been a minimum of 14 DMUs. Since this is only for illustrative purposes the lack of degrees of freedom is not of concern. However, note that there are means for overcoming this deficiency such as "window analyses" which are not discussed here.[6]

Finally, recall from the previous section that there are two ways to accommodate the "e" in the linear programming model. In this particular model we used the two-step process previously described.

The solution to this model is shown in Table 3. A quick interpretation of these results is that Base B is at best 92.7% efficient as indicated by the h_0^* value. In obtaining this evaluation, Base B was compared to Bases D and G as indicated by b_D and b_G. In addition, we can compute the input and output levels that Base B should attain in order to be rated 100% efficient via equations (3) and (4). These computations are shown in Table 4.

[Place Table 3 here.]

[Place Table 4 here.]

Interface With Cost Estimating/Analysis

How might DEA interface with cost analysis/estimating? In recent years there has been much discussion on estimating and analyzing operating and support (O&S) costs and this appears as being one area where the DEA methodology might be beneficial.

As previously noted, DEA has been used in efficiency measurement and cost variance analysis. This application is analogous to the use of standard cost accounting systems in private industry. With this application we are analyzing the level of inputs consumed (costs incurred) for the outputs produced. Did the organization consume too many resources and thereby incur too much expense in accomplishing its mission (producing output)? This type of use of DEA has been successfully illustrated by Bowlin [4] and Clark [5 and 10] for Air Force activities. As noted before, Bowlin analyzed real property maintenance activities in the Air Training Command and Tactical Air Command where he was able to identify not only which real property maintenance activities were operating inefficiently but also the sources and amounts of input inefficiencies. Clark reviewed the efficiency of aircraft maintenance activities in the Tactical Air Command.

However, there are other possible applications which still need to be explored. One possible future application is for this methodology to be used to compute efficient cost factors. One of

the outputs of DEA is efficient marginal rates of productivity and substitution. The marginal productivities can be used to compute cost factors for particular outputs which can be used in much the same way that regression coefficients are used. These cost factors can be used possibly in conjunction with goal programming or generalized networks for costing base operations for the programming and budgeting process or with various cost estimating accounting models. One problem which needs to be resolved before this particular application can be implemented is that those variables which have slack values (indicated by s_i^- and s_r^+ in model (2)) in the solution have no marginal productivities nor marginal rates of substitution from DEA.

Another possible use of DEA output is in estimating the O&S cost for a particular base. DEA would determine the efficient output and input (cost) levels. These efficient points could then be used in a regression model for predicting efficient cost levels given certain base characteristics or output. This would provide a "should" cost estimate.

Finally, DEA is a form of benefit/cost analysis where the outputs are the benefits and the inputs are the cost. Therefore, it might be used in a traditional benefit/cost analysis in determining whether to proceed with an investment decision or ranking investment/budget programs. In addition, since DEA handles multiple outputs and multiple inputs simultaneously, this opens the door for a benefit/cost analysis that does not require all benefits and costs to be reduced to dollar values.

It is recognized that the description of how DEA might interface with cost analysis/estimating has been somewhat vague. However, this is mainly due to these applications being untested and requiring further research into their applicability.

Conclusion

In this article we introduced an analytical technique developed by Charnes, Cooper, and Rhodes called data envelopment analysis. We discussed some of its characteristics and presented an example on how it is implemented and interpreted. As with any analytical technique or model, users should be completely familiar with the attributes and characteristics of the DEA model before he/she implements it. The example presented here was very basic and solely for the purpose of illustrating DEA. The complexities, intricacies, and importance of identifying the input/output variables and interpreting the results were not discussed, but should be completely reviewed before any application of DEA.

Even though DEA has been used only in a manner analogous to standard cost accounting systems, i.e., efficiency measurement, variance analysis, and controlling costs, it appears it has potential for application in more traditional cost analysis and estimating roles such as cost factor development, benefit/cost analysis, and O&S cost projections. However, research into this potential still needs to be accomplished.

TABLE 1

DESCRIPTION OF INPUT/OUTPUT MEASURES

INPUT MEASURES	DESCRIPTION
Supply Expenses (DOL)	Measures the availability of supplies and equipment with which to accomplish work.
Available Direct Labor Hours (LAB HR)	Represents the amount of time the work force is available for accomplishing civil engineering work.
Available Passenger Carrying Vehicles (VEH)	Measures number of vehicles available for transporting workers to the job site.
OUTPUT MEASURES	
Completed Work Orders (CWO)	Measures major work resulting in capitalization of real property records.
Completed Job Orders (CJO)	Represent day-to-day maintenance and repair work.
Completed Recurring Work Actions (CRWA)	Include recurring (preventive) maintenance and services for which the level of effort is known without an earlier visit to the job site.
Delinquent Job Order (DJO)	A job order not completed within the time specified by Air Force regulations. Checks the timeliness of work accomplishment which is essential to maintaining customer satisfaction.

TABLE 2

OBSERVED INPUT/OUTPUT VALUES

	Input Values			Output Values			
Base	VEH	DOL($000)	LAB HR	CWO	CJO	CRWA	DJO
A	44	$2,444.7	265,866	197	16,878	12,860	2,405
B	42	2,887.3	338,611	135	30,130	3,492	4,951
C	105	4,304.1	526,896	162	29,690	11,361	21,806
D	44	2,302.4	237,136	327	30,110	7,075	3,523
E	19	1,787.8	210,869	193	12,348	9,244	2,465
F	71	3,823.8	254,816	122	15,593	5,981	1,387
G	14	2,127.6	176,146	579	17,060	9,756	3,724
H	35	3,668.1	306,498	190	14,800	10,546	10,464
I	38	2,406.9	237,951	171	10,773	8,453	1,928

Key:

VEH - Available Passenger Carrying Vehicles

DOL - Supply Expenses

LAB HR - Available Direct Labor Hours

CWO - Completed Work Orders

CJO - Completed Job Orders

CRWA - Completed Recurring Work Actions

DJO - Delinquent Job Orders

TABLE 3

DEA SOLUTION FOR BASE B

$h_0^* = T^* = .927$

$s_{CWO}^{+*} = 373$

$s_{CJO}^{+*} = 0$

$s_{CRWA}^{+*} = 6,226$

$s_{DJO}^{+*} = .00013$

$s_{VEH}^{-*} = 0$

$s_{DOL}^{-*} = 0$

$s_{LABHR}^{-*} = 58,035$

$b_D^* > 0$

$b_G^* > 0$

TABLE 4

EFFICIENT INPUT/OUTPUT LEVELS FOR BASE B

Outputs

$\hat{y}_r = y_r + s_r^{+*}$ (Equation 3)

$\hat{y}_{CWO} = 135 + 373 = 508$

$\hat{y}_{CJO} = 30,130 + 0 = 30,130$

$\hat{y}_{CRWA} = 3,492 + 6,226 = 9,718$

$\hat{y}_{DJO} = .0002 + .00013 = .00033; \quad 1/.00033 = 3,030$

Inputs

$\hat{x}_i = h_0^*(x_i) - s_i^{-*}$ (Equation 4)

$\hat{x}_{VEH} = .927(42) - 0 = 38.9$

$\hat{x}_{DOL} = .927(2,887.3) - 0 = 2,676.5$

$\hat{x}_{LABHR} = .927(338,611) - 58,035 = 255,857$

FOOTNOTES

1. This term is used to indicate a not-for-profit entity in which a manager has some freedom of decision making on its inputs and outputs. E.g., squadron, wing, base, directorate, etc.

2. These are referred to as "virtual multiplier" or "transformation rates" by Charnes, Cooper, and Rhodes [7] and their values are determined from the data. they should not be confused with any type of subjective weighting scheme.

3. See footnote 2.

4. An asterisk (*) indicates that the variable value is the optimal value resulting from the solution of the DEA model.

5. The functions d(O-D')/d(O-D) and d(O-E')/d(O-E) are to be understood as measures of distance from the origin in Euclidean metric.

6. For further discussion on window analyses see Bowlin [4] and Charnes, Clark, Cooper, and Golany [5].

REFERENCES

1. Banker, R.D.; Charnes, A.; and Cooper, W.W. "Estimation of Technical and Scale Inefficiencies in Data Envelopment Analysis." Management Science (1984).

2. Banker, R.D. and Morey, R. "Data Envelopment Analysis with Categorical Inputs and Outputs." Management Science (submitted).

3. Banker, R.D. and Morey, R. "Efficiency Analysis for Exogeneously Fixed Inputs and Outputs." Operations Research (submitted).

4. Bowlin, W.F. "A Data Envelopment Analysis Approach To Performance Evaluation In Not-For-Profit Entities With An Illustrative Application To The U.S. Air Force." Ph.D. dissertation, The University of Texas at Austin, Texas (December 1984).

5. Charnes, A.; Clark, C.T.; Cooper, W.W.; and Golany, B. "A Developmental Study Of Data Envelopment Analysis In Measuring The Efficiency Of Maintenance Units In The U.S. Air Forces." The Annals of Operations Research (forthcoming).

6. Charnes, A. and Cooper, W.W. "Programming With Linear Fractional Functionals." Naval Research Logistics Quarterly 9, 3-4 (1962).

7. Charnes, A.; Cooper, W.W.; and Rhodes, E. "Measuring the Efficiency of Decision Making Units." European Journal of Operational Research 2, 6 (November 1978): pp. 429-44.

8. Charnes, A.; Cooper, W.W.; and Rhodes, E. "Short Communication: Measuring the Efficiency of Decision Making Units." European Journal of Operational Research 3, 4 (July 1979): p. 339.

9. Charnes, A.; Cooper, W.W.; and Thrall, R. "An Archimedean Approach to CCR Ratio Efficiency Analysis." Paper presented to the ORSA/TIMS Joint National Meeting (28 November 1984).

10. Clark, C.T. "Data Envelopment Analysis And Extensions For Decision Support And Management Planning." Ph.D. dissertation, The University of Texas at Austin, Texas (May 1983).

END

FILMED

1-86

DTIC